

Research Article

Context-Aware AAL Services through a 3D Sensor-Based Platform

Alessandro Leone, Giovanni Diraco, and Pietro Siciliano

Institute for Microelectronics and Microsystems, Italian National Research Council (CNR), Via Monteroni, c/o Campus Università del Salento, Palazzina A3, Lecce, Italy

Correspondence should be addressed to Alessandro Leone; alessandro.leone@le.imm.cnr.it

Received 8 February 2013; Accepted 27 April 2013

Academic Editor: Eugenio Martinelli

Copyright © 2013 Alessandro Leone et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The main goal of Ambient Assisted Living solutions is to provide assistive technologies and services in smart environments allowing elderly people to have high quality of life. Since 3D sensing technologies are increasingly investigated as monitoring solution able to outperform traditional approaches, in this work a noninvasive monitoring platform based on 3D sensors is presented providing a wide-range solution suitable in several assisted living scenarios. Detector nodes are managed by low-power embedded PCs in order to process 3D streams and extract postural features related to person's activities. The feature level of details is tuned in accordance with the current context in order to save bandwidth and computational resources. The platform architecture is conceived as a modular system suitable to be integrated into third-party middleware to provide monitoring functionalities in several scenarios. The event detection capabilities were validated by using both synthetic and real datasets collected in controlled and real-home environments. Results show the soundness of the presented solution to adapt to different application requirements, by correctly detecting events related to four relevant AAL services.

1. Introduction

During the last years, the interest of the scientific community in smart environments has grown very fast especially within the European Ambient Assisted Living (AAL) Program with the aim of increasing the independent living and the quality of life of older people. The design of AAL systems is normally based on the use of monitoring infrastructures provided by smart environments. Such infrastructures include heterogeneous sensing devices in ad hoc networks with distributed data processing resources and coordinated by intelligent agents, offering information analysis and decision making capabilities. Human activities monitoring is a crucial function of AAL systems, especially in detection of critical situations (e.g., falls and abnormal behaviors) or as support during the execution of relevant tasks (e.g., daily living activities and training/rehabilitation exercises). Generally, human monitoring systems are based on both wearable devices or environmental equipment. In the first case, markers or kinematic sensors (e.g., MEMS accelerometers or

gyroscopes) are worn by the end user for body's movements detection. Recently Baek et al. [1] have presented a necklace embedding a triaxial accelerometer and a gyroscope able to distinguish falls from regular ADLs (activities of daily living) by measuring the angle of the upper body to the ground. Although wearable techniques could be accurate and suitable even in outdoor conditions, their efficiency is limited due to the false alarm occurrences if the device is incorrectly worn or not worn at all (after having a bath or during the night), as described by Kröse et al. [2]. On the other hand, ambient devices are normally based on some kind of vision system (monocular or stereo cameras, infrared camera, 3D sensors, etc.) for people tracking or sensors installed in household furniture/appliances (pressure sensors on the floor, on/off switches, etc.) inferring the person's activities. Rimminen et al. [3] have been suggested an innovative floor sensor able to detect falls by using the near-field imaging (NFI) principle. However, ambient sensors require typically an ad hoc design or redesign of the home environment. The vision-based techniques are the most affordable and accurate solutions due to

the richness of acquired data and the installation accessibility that do not require an ad hoc redesign of the environment. Foroughi et al. [4] have been presented a passive vision (monocular camera) approach for monitoring human activities with a particular interest in the problem of fall detection. More recently, Edgcomb and Vahid [5] have made an attempt to detect falls on privacy-enhanced monocular video. On the other hand, the visions systems based on 3D sensors of new generation of both TOF (Time of Flight) and non-TOF (e.g., structured light) overcome typical issues affecting the passive vision systems such as the dependence on ambient conditions (e.g., brightness, shadows, and chromaticity and appearance of surfaces) and the poor preservation of the privacy. Indeed, 3D sensors enable a new visual monitoring modality based on the metric reconstruction of a scene by processing only distance maps (i.e., range images), guaranteeing person's privacy. The use of range images simplifies both preprocessing and feature extraction steps, allowing the use of less computational expensive algorithms more suitable for embedded PCs typically installed in AAL contexts. The usage of cheaper (non-TOF) 3D sensors have been recently described by Mastorakis and Makris [6] for the detection of falls in elderly people. However non-TOF sensors estimate distances from the distortion of an infrared light pattern projected on the scene, so their accuracy and covered range are seriously limited (distance up to 3-4 m with 4 cm accuracy) as reported by Khoshelham [7]. However, such sensors are designed especially for human motion capture in entertainment/gaming applications and then optimized for nonoccluded frontal viewing. TOF sensors, employing the so-called Laser Imaging Detection and Ranging (LIDAR) technique, estimate distances more reliably and accurately by measuring the delay between emitted and reflected light, so they can reach longer distances within a wider Field-of-View (FOV). Mixed approaches based on different kind of sensors are also possible, in which data coming from heterogeneous sensors are correlated (data fusion) to compensate false alarms, making the solution more reliable. A multisensor system for fall detection using both wearable accelerometers and 3D sensors has been described by Grassi et al. [8].

AAL systems present other critical issues concerning interoperability, modularity, and hardware independence, as observed by Fuxreiter et al. [9]; devices and applications are often isolated or proprietary, preventing the effective customization and reuse causing high development costs and limiting the combination of services in order to adapt to user's needs. New trends are addressed to integrate an open middleware in AAL systems, as a flexible intermediate layer able to accommodate requirements and scenarios. Relevant AAL-oriented middleware architectures have been described by Wolf et al. [10], Schäfer [11], and Coronato et al. [12].

This paper presents a novel monitoring platform based on 3D sensors for AAL services delivering in smart environments. The proposed solution is able to process data acquired by different kind of 3D sensors and it is suitable to be integrated in AAL-oriented middleware providing monitoring functionalities in several AAL applications.

2. Materials and Methods

The platform (architecture) is conceived as modular, distributed, and open, by integrating several detector nodes and a coordinator node (Figure 1). It is designed to be integrated into wide AAL systems through open middleware. Three main logical layers have been defined: data processing resource, sensing resource, and AAL services management. The data processing resource layer is implemented by both detector nodes (Figure 1(a)) and coordinator node (Figure 1(b)). Moreover, the detector nodes implement the sensing resource layer. The 3D sensor network has a hierarchical topology as shown in Figure 2(a), composed of M detector nodes managing several 3D sensor nodes for each and one coordinator node that receives high-level reports from detector nodes. Both the 3D sensor, shown in Figure 2(b), and the embedded PC implementing the coordinator and detector nodes, shown in Figure 2(c), are low-power, compact, and noiseless devices, in order to meet typical requirements of AAL contexts.

A detector node can handle either overlapping 3D views (i.e., frames captured by distinct 3D sensors having at least a few common points) or nonoverlapping ones, whereas 3D views managed by distinct detector nodes must be always nonoverlapped. 3D data streams are fused at level of single detector node, whereas high-level data are fused at both detector and coordinator nodes. The detector nodes are responsible for events involving either single view or overlapping views, using data fusion to resolve occlusions. Instead, the coordinator handles events involving nonoverlapping views (inter-view events) and it is responsible for achieving a global picture of the events (e.g., the detection of a wandering state with a recovered fall in the bedroom and an unrecovered fall in the kitchen). Since AAL systems are typically implemented to assist elderly people living alone, the issue of inter-view people identification has not been addressed (i.e., only one person at a time is assumed present in the home).

The coordinator layer includes the architectural modules for the management of the detector nodes (control and data gathering), high-level data fusion, inter-view event detection and context management. The system manager (Figure 1(c)) managing the whole AAL system that includes the monitoring platform presented in this work as a functional component. It is inspired to the open AAL middleware UniversAAL [13], in order to achieve global AAL services' goals. Each aforementioned architectural layer will be described in detail in the following.

2.1. The Sensing Resource. The sensing nodes are the 3D sensors connected to each detector node. Figure 3 shows one sensing node in a wall-mounted configuration with its extrinsic calibration parameters (θ : tilt angle, β : roll angle, H : height with respect to the floor plane) referring to the sensor reference system O_S . The figure shows also the world reference system O_W fixed on the floor plane. The 3D sensors management module is devoted to read out from the 3D sensor the data stream (three Cartesian matrices X , Y , and Z in O_S coordinates). The preprocessing module includes functionalities for extrinsic calibration and people detection

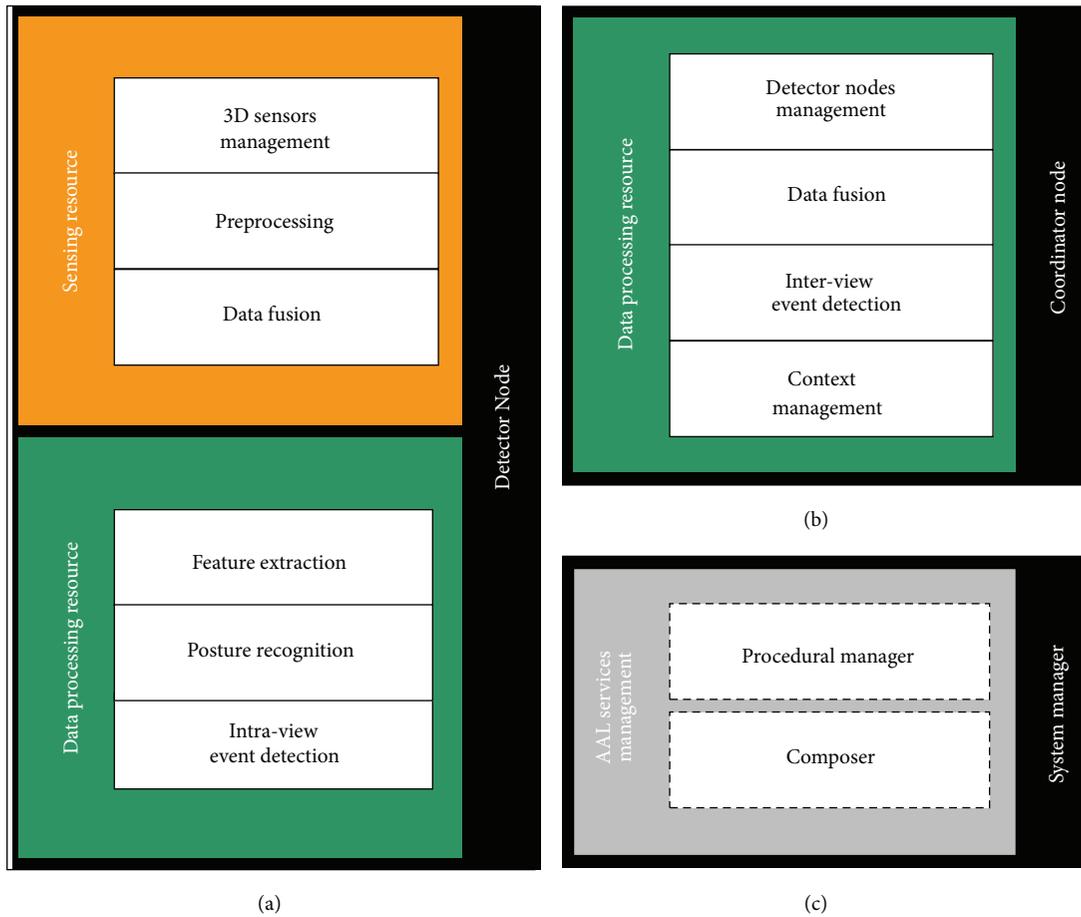


FIGURE 1: Overall architecture representation. (a) Detector node, (b) coordinator node, (c) AAL system manager.

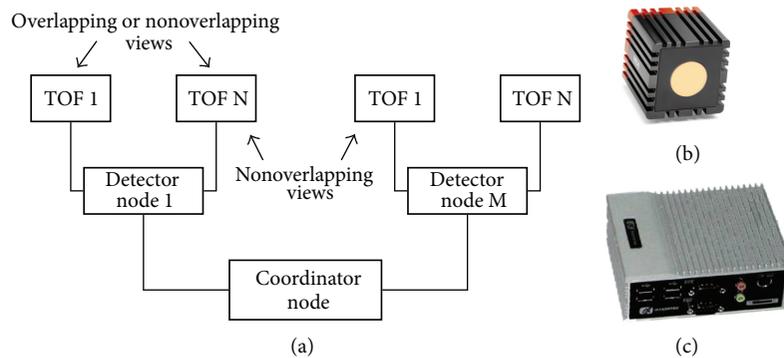


FIGURE 2: (a) Hierarchical network topology; (b) 3D TOF sensor (MESA SR-4000) adopted in this study during experimentation; (c) Intel Atom processor-based embedded PC used for both detectors and coordinator nodes.

in points cloud. The extrinsic calibration is fully automated (self-calibration) to simplify the sensor installation without requiring neither calibration tool nor user intervention. The self-calibration procedure is based on the RANSAC floor plane detection, suggested by Gallo et al. [14], in order to estimate the reference system change from O_S to O_W in which the feature extraction process is done. To identify a person in the acquired data, a set of well-known vision

processing steps, namely, background modeling, segmentation, and people tracking, is implemented according to a previous authors' study [15]. The Mixture of Gaussian dynamical model, proposed by Stauffer and Grimson [16], is used for background modeling since it is able to rapidly compensate little variations in the scene (e.g., movements of chairs and door opening/closing). For person detection and segmentation, a specific Bayesian formulation is defined in

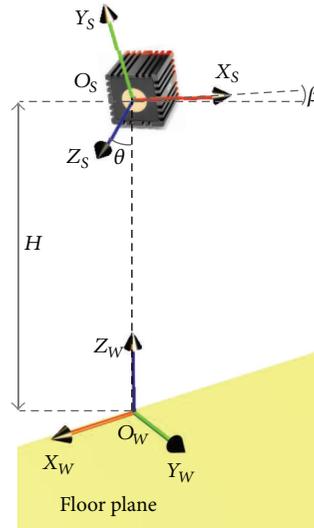


FIGURE 3: Wall-mounted TOF sensor with extrinsic calibration parameters (H, θ, β) and two reference systems: sensor fixed (O_S) and floor plain fixed (O_W).

order to filter out nonperson objects even in very cluttered scenes.

A multiple hypothesis tracking, by using the conditional density propagation overtime, proposed by Isard and Blake [17], is implemented to track people effectively even in presence of large semioccluding objects (e.g., tables and sofas) as frequently happen in home environments. Considering that only one person at a time can be present in the home, the range data generated from overlapping 3D views are fuse together by using a triangulation-based prealignment refined with the fast Iterative Closest Point as suggested by Won-Seok et al. [18]. Finally, a middleware module plugs in 3D sensors into the whole system providing also a semantic description of the 3D data. This module is able to handle different type of 3D sensors, both TOF and non-TOF, translating the specific data format into an abstract one common for all 3D sensors plugged in the system.

2.2. The Data Processing Resource. Two different kinds of data processing resources are defined, namely, the detector and the coordinator, of which details are provided in this section.

2.2.1. The Detector. The detector data processing resource includes the following modules: feature extraction, person's position, posture recognition, and intra-view event detection. Features are extracted from 3D data by using two body descriptors having different level of details and computational complexity. Coarse grained features are extracted by using a volumetric descriptor that exploits the spatial distribution of 3D points represented in cylindrical coordinates (h, r, θ) corresponding to height, radius, and angular locations, respectively. The rotational invariance is obtained by choosing the h -axis related to the body's vertical central axis and suppressing the angular dimension θ . Instead, the scale invariance obtained normalizing by the size of the reference cylinder, whereas the translational invariance is guaranteed

by placing the cylinder axis on the body's centroid M . Thus the 3D points are grouped into rings orthogonal to and centered at the h -axis while sharing the same height and radius (Figure 4(a)), showing the cylindrical reference system with highlighting the k th ring and its included 3D points. The corresponding volumetric features are represented by the cylindrical histogram shown in Figure 4(b) obtained by taking the sum of the bin values for each ring. The fine grained features are achieved by using a topological representation of body information embedded into the 3D point cloud. The intrinsic topology of a generic shape, that is, a human body scan captured via TOF sensors, is graphically encoded by using the discrete Reeb graph (DRG) proposed by Xiao et al. [19]. To extract the DRG, the geodesic distance is used as invariant Morse function [20] since it is invariant not only to translation, scale, and rotation but also to isometric transformations ensuring the high accuracy of the body parts representation under postural changes. The geodesic distance map is computed by using a two-step procedure. At first, a connected mesh is built on the 3D point cloud (Figure 4(c)) by using the nearest-neighbor rule. Then, assuming M the starting point (i.e., the body's centroid), the geodesic distance between M and each other mesh node is computed as the shortest path on mesh by using an efficient implementation of Dijkstra's algorithm [21]. The computed geodesic map is shown in Figure 4(d) in which false colors represent geodesic distances. The DRG is extracted by subdividing the geodesic map in regular level sets and connecting them on the basis of an adjacency criterion as described by Diraco et al. [22] suggesting also a methodology to handle self-occlusions (i.e., due to body parts overlapping other body parts). The DRG-based features are shown in Figure 4(e) and represented by the topological descriptor that includes DRG nodes $\{\dots, C_i, C_j, C_k, \dots\}$ and related angles θ_{ijk} . The person's position is defined in terms of 3D coordinates with respect to the world reference systems O_W associated with the 3D sensor in case of single (nonoverlapping) view.

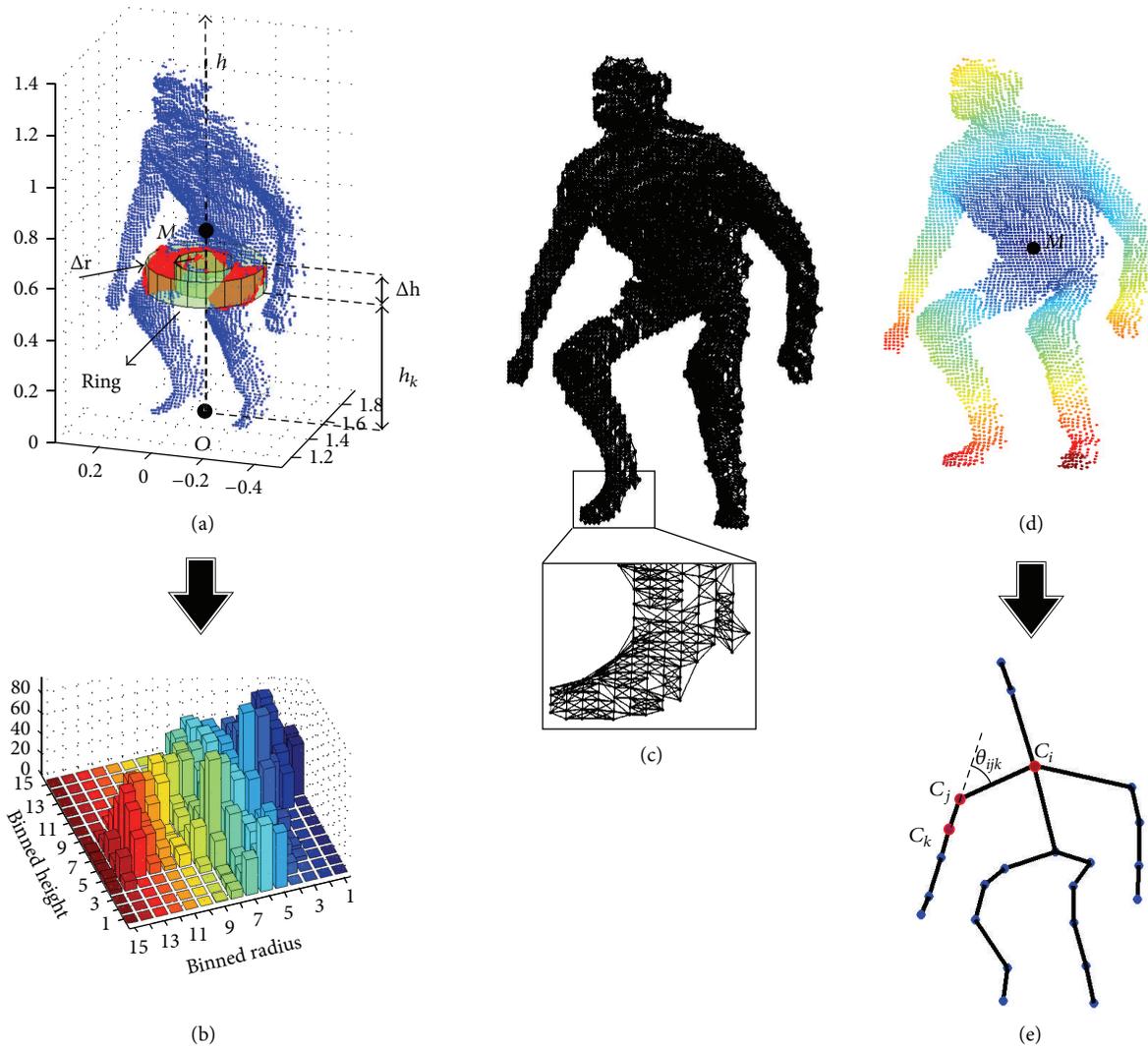


FIGURE 4: (a) Volumetric representation of body 3D point cloud. (b) Cylindrical histogram features extracted from the volumetric representation. (c) Connected mesh. (d) Geodesic map in false colors used to represent distances: near points are blue, far red. (e) Reeb graph features extracted from geodesic map.

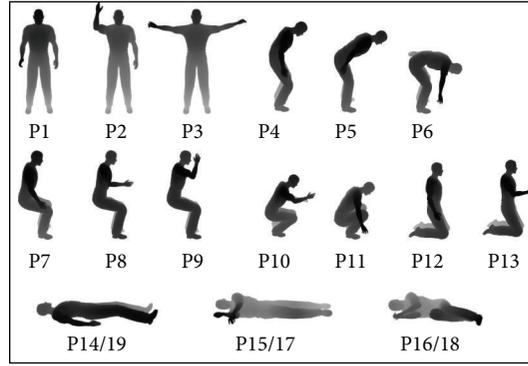
In case of overlapping views the 3D position is estimated via triangulation (i.e., considering at least two sensor views and the relative position of the person in them) and assuming one of the overlapping views as reference view. In activity monitoring and related fields, the body posture is considered a crucial element, since a typical human action can be decomposed in few relevant key postures differing significantly from each other and suitable to both representing and inferring activities and behaviors, as pointed out by Brendel and Todorovic [23] and by Cucchiara et al. [24], respectively. To cover as wide range as possible of AAL applications, a large class of key postures organized into four levels with different details has been defined as described in the following. The considered levels are summarized in Figure 5(a). At the first level, the four basic postures, Standing (St), Bending (Be), Sitting (Si), and Lying down (Ly), are extracted. At the second level, the person's centroid height with respect to the floor plane is taken into account in order to discriminate, for

instance, a “Lying down on bed” from a “Lying down on floor.” The orientation of the body's torso is taken into account by the third level. The fourth and final level describes the body's extremities configuration, providing a total amount of 19 postures. A sample TOF frame for each kind of posture is shown in Figure 5(b). Given the coarse-to-fine features extracted as previously discussed, the multiclass formulation of the SVM (Support Vector Machine) based on the one-against-one strategy described by Debnath et al. [25] is used for postures classification. Since interesting results have been obtained when Radial Basis Function (RBF) kernel is used [26], key kernel parameters (regularization constant K and kernel argument γ) are adjusted according to a global grid search strategy.

Furthermore, the detector node is responsible to detect events related to both nonoverlapping and overlapping views (such as falls or activities happening entirely into the same detection area). In general, human actions are recognized

Level 1	Level 2	Level 3	Level 4	P		
Standing (St)			Arms along body	1		
			One arm up	2		
			T shape	3		
Bending (Be)		Slightly bowed	Arms along the body	4		
		Medium bowed	Arms along the body	5		
		Fully bowed	One arm picking up an object	6		
Sitting (Si)	On chair		Arms along body	7		
			One arm ahead	8		
			One arm up	9		
	On floor	Crouching		One arm ahead	10	
				One arm picking up an object	11	
				Arms along the body	12	
		Kneeling			One arm ahead	13
					Spread legs	14
		Lying down (Ly)	On floor		Spread legs	15
					Curled up legs	16
Spread legs	17					
On bed				Curled up legs	18	
				Spread legs	19	

(a)



(b)

FIGURE 5: (a) Body posture classes organized in four different levels of detail and (b) corresponding sample TOF frames.

by considering successive postures over a time period. A transition action occurs when the person changes the current action to another action. Thus, a transition action might include several transition postures. Such transition postures are recognized at detector node level by using a backward search strategy, whereas transition actions are recognized by the coordinator. Starting from the current 3D frame, the previous N frames are checked if they refer to similar postures. In this case, the transition action is recognized, otherwise the backward search strategy continues with another 3D frame. Recognized transition postures are sent to the coordinator responsible for events detection in nonoverlapping views. If transition postures yield a meaningful event (e.g., fall) it is also sent to the coordinator. Finally, the detector data processing is plugged in via middleware module as data processing resource able to process data coming from 3D sensor resources and to communicate with the coordinator.

2.2.2. The Coordinator. The coordinator data processing resource includes the following functional modules: detector nodes management, data fusion, inter-view event detection, and context management. Information concerning detector nodes gathered by the coordinator includes the node position within the home (e.g., living room and bedroom) and the adjacency of node views (i.e., if two nonoverlapping views are directly accessible or if they are accessible through a third view). In addition, on the basis of the current application context the detector nodes are configured according to the most appropriate level of details and the events of interest. The data reports gathered from detectors are fused together on the basis of both detector position and timestamp. The interview events are detected by using a backward search strategy similar to that already described in the previous subsection but recognizing transition actions instead of transition postures. The transition actions are merged together to form single atomic actions whereas global events are recognized by using Dynamic Bayesian Networks (DBNs) specifically designed for each application scenario, following an approach similar to the one proposed by Park and Kautz [27]. The designed DBNs have a hierarchical structure with three node layers named activity, interaction, and sensor. The activity

layer stays on top of the hierarchy and includes hidden nodes to model high-level activities (i.e., ADLs, behaviors, etc.). The interaction layer is an hidden layer as well and it is devoted to model the states of evidence for interactions inside the home (i.e., appliances and furniture locations, and person's position). The sensor layer, at the bottom of hierarchy, gathers data from detector sensors: locations and postures. Each DBN is hence decomposed in multiple Hidden Markov Models (HMMs) including interaction and sensor layers and trained on the basis of the Viterbi algorithm as described by Jensen and Nielsen [28].

2.3. System Manager. The system manager refers to the whole AAL system management by means of the definition and the execution of procedures and workflows which react to situations of interest. While such situations are identified by the context manager at charge of the coordinator, the system manager through the procedural manager handles the workflow on the basis of what the system is required to react (e.g., sending an alarm message). Furthermore, the procedural manager outlines service goals in an abstract way, whereas the composer is responsible for combining available AAL services to achieve such goals.

3. Results and Discussion

Invariance properties and recognition performance of suggested coarse-to-fine features were assessed by using realistic synthetic postures generated as suggested by Gond et al. [29]. A large posture datasets of about 6840 samples, with and without semiocclusions, was generated under different angles (from 0° to 180° with 45° steps) and distances (Low: 2 m, Med: 6 m, High: 9 m). The 4-quadrant technique suggested by Li et al. [30] was adopted to simulate semiocclusions similar to those normally present in home environments. The achieved recognition rates detailed for each feature level, angle, and distance are reported in Table 1.

At levels 1 and 2, the volumetric descriptor exhibited a good recognition rate that without semiocclusions was in average equal to 96% (average taken over all angles and distances), decreasing to 93% in presence of semiocclusions.

TABLE 1: Classification rates (%) of postures at varying viewpoint angles and distances, both with and without semiocclusions.

Feature	Level 1			Level 2			Level 3			Level 4				
	Distance	Low	Med	High	Low	Med	High	Low	Med	High	Low	Med		High
Volumetric descriptor														
View angle (°)	0	99	96	94	98	97	96	95	94	90	74	73	73	Without semi-occlusions
	45	99	97	95	97	96	94	93	93	89	73	71	72	
	90	98	97	93	98	97	93	94	93	88	75	74	70	
	135	97	95	96	97	94	95	92	92	91	77	75	72	
	180	99	96	95	96	95	96	95	92	88	74	74	74	
	0	94	95	92	94	92	93	91	88	87	67	66	67	With semi-occlusions
	45	94	93	93	94	94	92	93	87	86	66	64	64	
	90	95	94	94	91	90	91	89	86	86	67	65	65	
	135	93	90	92	93	94	92	84	85	85	65	67	67	
	180	94	92	91	92	91	93	90	85	84	68	64	66	
Topological descriptor														
View angle (°)	0	99	95	94	98	96	93	97	93	87	98	91	86	Without semi-occlusions
	45	98	95	93	97	94	91	96	92	86	99	85	87	
	90	94	91	90	93	92	89	94	87	79	93	84	79	
	135	98	97	95	96	93	92	97	90	86	95	87	83	
	180	99	98	94	95	94	94	98	95	88	97	89	86	
	0	94	89	81	93	86	82	92	87	77	91	87	80	With semi-occlusions
	45	91	85	84	90	85	79	94	84	75	92	83	79	
	90	87	78	69	88	75	73	86	79	68	86	78	75	
	135	92	86	75	87	83	75	89	76	71	88	81	78	
	180	90	84	77	91	80	78	93	82	78	90	82	81	

The topological descriptor (levels 1 and 2) exhibited a classification rate without semiocclusions in average equal to 95% and 94%, respectively, demoting to 84% and 83%, respectively, with semiocclusions. Results suggest that the volumetric descriptor is more robust to semiocclusions than the topological one. At level 3 the volumetric descriptor showed an acceptable classification rate on average of 92% that demoted to 87% with semiocclusions. At this level the topological descriptor gave an average classification rates of 91% and 82% without and with semiocclusions, respectively. At level 4 the two descriptors exhibited the major differences. In fact, the volumetric descriptor achieved very poor classification rates, whereas the topological descriptor was able to discriminate well high-level postures achieving without semiocclusions and at Low distances an average of 96%, and of 89% at all distances (up to 10 m), decreasing to 83% in presence of semiocclusions. Whereas the volumetric classification rate was almost uniform across angles and distances, the topological one tended to decrease with distance and in correspondence to viewpoint angles of 90° (lateral position) at which self-occlusions were most heavy.

The event detection performance was evaluated in real-home environments by involving ten healthy subjects, 5 males, and 5 females, having different physical characteristics: age 31 ± 6 years, height 173 ± 20 cm, and weight 75 ± 22 kg.

Figure 6(a) shows sixteen sample 3D frames of the collected dataset. The typical apartment is shown in Figure 6(b) with the locations (from 1 to 11) in which actions have been performed. The sensor network used during experiments included three sensor nodes, S1 in the bedroom, and S2 and S3 in the living room with overlapping views. Each sensing node is represented by the MESA SwissRanger SR-4000 [31] shown in Figure 1(b), that is, a state-of-the-art TOF 3D sensor with compact dimensions ($65 \times 65 \times 68$ mm), noiseless functioning (0 dB noise), QCIF resolution (176×144 pixels), long distance range (up to 10 m), and wide ($69^\circ \times 56^\circ$) FOV (Field-Of-View). Since the SR-4000 sensor comes intrinsically calibrated by the manufacturer, the calibration procedure had to estimate only the extrinsic calibration parameters. The 3D sensors were managed by two logical detector nodes: one for S1 and another one for both S2 and S3. The two logical detectors and the coordinator were implemented into the same physical node which was an Intel Atom 1.6 GHz processor-based embedded PC shown in Figure 1(c). Four relevant AAL services have been considered, namely, fall detection, wandering detection, ADLs recognition, and training exercises recognition. One dataset for each service was collected and characterized by different combinations of occlusions, distances, angles, and feature levels as reported in Table 2 by columns from 2 to 8.

TABLE 2: Detection performance of four event classes.

Detected event	Semi-occlusions	Angles	Distances	Feature level frequency (%)				Detection performance (%)	
				Lev. 1	Lev. 2	Lev. 3	Lev. 4	TNR	TPR
Falls	Yes	All	All	90	10	0	0	97.5	83.0
Wandering	Yes	All	All	50	20	30	0	92.7	81.6
Activities of daily living	Few	All	Med	50	20	25	5	98.3	96.4
Training exercises	No	Front	Low	5	5	40	50	99.2	95.6

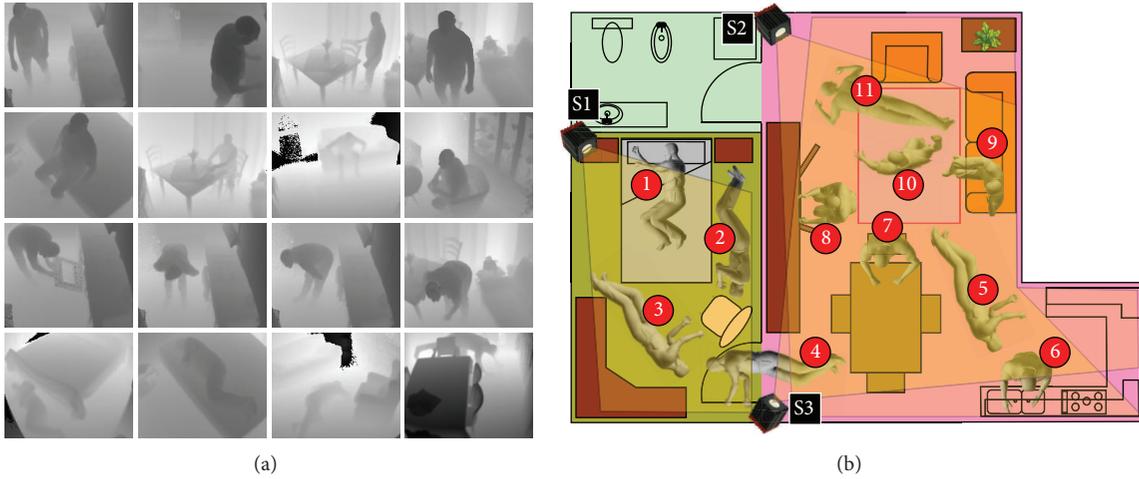


FIGURE 6: (a) Sample TOF frames collected in real homes. (b) Map of apartment used for data collection.

The last two columns in Table 2 report the achieved detection performance in terms of True Negative Rate (TNR) and True Positive Rate (TPR) measures defined as follows:

$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad \text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (1)$$

where TP, TN, FP, and FN stand for True Positive, True Negative, False Positive, and False Negative, respectively.

Concerning the fall detection scenario, different fall (locations 2, 3, 4, and 5 in Figure 6(b)) and nonfall (locations 1 and 11 in Figure 6(b)) actions were performed, with and without the presence of occluding objects, in order to evaluate discrimination performance. The system was able to discriminate correctly falls even in presence of semiocclusions, achieving 97.5% and 83% of TNR and TPR, respectively. Since fall events were detected at level of detector node (intra-view), ambiguous situations such as those in which a fall was located between nonoverlapping views (e.g., location 4 in Figure 6(b)) gave rise to False Negatives. The wandering state, instead, was detected at coordinator level since it normally involves several nonoverlapping views. In general, it is not simple to detect a wandering state since it is not just an aimless movement. Rather, it is a “purposeful behavior initiated by a cognitively impaired and disoriented individual characterized by excessive ambulation” as stated by Ellelt [32]. On the basis of such characterization, wandering states were discriminated from ADLs with 92.7% and 81.6% of TNR and TPR, respectively. While for fall detection the involved feature detail levels were almost exclusively the first two

with prevalent adoption of the volumetric representation, in the case of wandering detection also the third feature level was involved with a moderate topological representation. The following seven kinds of activities were performed in order to evaluate the ADLs recognition capability: sleeping, waking up, eating, cooking, housekeeping, sitting to watch TV, and physical training. In this case all four feature levels were involved, although the fourth level had a low incidence (5%). ADLs were recognized with 98.3% and 96.4% of TNR and TPR, respectively. A moderate misclassification was observed for housekeeping activities since they occasionally were erroneously recognized as wandering state. For the training exercises scenario, a virtual trainer was developed instructing participants to follow a physical activity program and to perform the recommended exercises correctly. The recommended physical exercises were of the following kinds: biceps curl, squatting, torso bending, and so forth. Involved feature levels were, for the majority, the last two (40% and 50%, resp.), with prevalent use of topology-based features. The performed exercises were correctly recognized achieving 99.2% and 95.6% of TNR and TPR, respectively. The detection results reported in Table 2 show that the system was able to select the most appropriate level of feature details almost in all scenarios. The most computationally expensive steps were preprocessing, feature extraction, and posture classification. They were evaluated in terms of processing time that was constant for preprocessing and classification resulting, respectively, in 20 ms and 15 ms per frame. The volumetric descriptor has taken an average processing time of about

TABLE 3: Detection performance of four related studies.

Related studies	Devices	TNR (%)	TPR (%)
Baek et al. [1]	Tri-axial accelerometer + gyroscope sensors	100.0	81.6
Rimminen et al. [3]	Near-field imaging floor sensor	91.0	91.0
Foroughi et al. [4]	Monocular passive vision	98.0	93.0
Edgcomb and Vahid [5]	Privacy-enhanced monocular passive vision	92.0	91.0

20 ms, corresponding to about 18 fps (frame per second). The topological approach, on the other hand, required a slightly increasing processing time among hierarchical levels from an average value of 40 ms to 44 ms due to the incremental occurrence of self-occlusions, achieving up to 13 fps.

Comparison with Related Studies. Different studies based on both wearable and ambient devices have been considered in order to compare the discussed results. All reported studies have been carried out by detecting abnormal behaviors (e.g., falls and wandering) among normal human activities (e.g., ADLs and physical exercises) in real or near-real conditions. Studies based on 3D sensors have not been reported yet; to from the author's knowledge, comprehensive works on abnormal behaviors detection cannot be found in the literature. The results of the related studies are reported in Table 3.

4. Conclusions

The main contribution of the work is the design and the evaluation of a unified solution for 3D sensor-based in-home monitoring for different context-aware AAL services. A modular platform has been presented, which is able to classify a large class of postures and detect events of interest, accommodating simple wall-mounting sensor installations. The platform was optimized and validated for embedded processing in order to meet typical AAL in-home requirements, such as low-power consumption, noiselessness, and compactness. The experimental results have shown that the system was able to effectively adapt to four different important AAL scenarios exploiting a context-aware multilevel feature extraction. The process guarantees a reliable detection of relevant events overcoming well-known problems affecting traditional vision-based monitoring systems in a privacy preserving way. The ongoing work concerns the on-field validation of the system that will be deployed in elderly apartments in support of two different AAL scenarios concerning the detection of dangerous events and abnormal behaviors.

Acknowledgment

The presented work has been carried out within the BAITAH project funded by the Italian Ministry of Education, University and Research (MIUR).

References

- [1] W. S. Baek, D. M. Kim, F. Bashir, and J. Y. Pyun, "Real life applicable fall detection system based on wireless body area network," in *Proceedings of the IEEE Consumer Communications and Networking Conference (CCNC '13)*, pp. 62–67, 2013.
- [2] B. J. A. Kröse, T. J. M. Oosterhout, and T. L. M. Kasteren, "Activity monitoring systems in health care," in *Computer Analysis of Human Behavior*, A. A. Salah and T. Gevers, Eds., pp. 325–346, Springer, London, UK, 2011.
- [3] H. Rimminen, J. Lindström, M. Linnavuo, and R. Sepponen, "Detection of falls among the elderly by a floor sensor using the electric near field," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 6, pp. 1475–1476, 2010.
- [4] H. Foroughi, B. S. Aski, and H. Pourreza, "Intelligent video surveillance for monitoring fall detection of elderly in home environments," in *Proceedings of the 11th International Conference on Computer and Information Technology (ICCIT '08)*, pp. 219–224, Khulna, Bangladesh, December 2008.
- [5] A. Edgcomb and F. Vahid, "Automated Fall Detection on Privacy-Enhanced Video," in *Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '12)*, pp. 252–255, San Diego, Calif, USA, 2012.
- [6] G. Mastorakis and D. Makris, "Fall detection system using Kinect's infrared sensor," *Journal of Real-Time Image Processing*, 2012.
- [7] K. Khoshelham, "Accuracy analysis of Kinect depth data," in *Proceedings of the ISPRS Workshop on Laser Scanning*, Calgary, Canada, August 2011.
- [8] M. Grassi, A. Lombardi, G. Rescio et al., "An integrated system for people fall-detection with data fusion capabilities based on 3D ToF camera and wireless accelerometer," in *Proceedings of the 9th IEEE Sensors Conference 2010 (SENSORS '10)*, pp. 1016–1019, Waikoloa, Hawaii, USA, November 2010.
- [9] T. Fuxreiter, C. Mayer, S. Hanke, M. Gira, M. Sili, and J. Kropf, "A modular platform for event recognition in smart homes," in *Proceedings of the 12th IEEE International Conference on e-Health Networking, Application and Services (Healthcom '10)*, July 2010.
- [10] P. Wolf, A. Schmidt, J. Parada Otte et al., "openAAL—the open source middleware for ambient-assisted living (AAL)," in *Proceedings of the AALLIANCE Conference*, Malaga, Spain, 2010.
- [11] J. Schäfer, "A middleware for self-organising distributed ambient assisted living applications," in *Proceedings of the Workshop Selbstorganisierende, Adaptive, Kontextsensitive Verteilte Systeme (SAKS '10)*, 2010.
- [12] A. Coronato, G. De Pietro, and G. Sannino, "Middleware services for pervasive monitoring elderly and ill people in smart environments," in *Proceedings of the 7th International Conference on Information Technology—New Generations (ITNG '10)*, pp. 810–815, Las Vegas, Nev, USA., April 2010.
- [13] UniversAAL project 2012, <http://www.universaal.org/>.
- [14] O. Gallo, R. Manduchi, and A. Rafii, "CC-RANSAC: fitting planes in the presence of multiple surfaces in range data," *Pattern Recognition Letters*, vol. 32, no. 3, pp. 403–410, 2011.
- [15] A. Leone, G. Diraco, and P. Siciliano, "Detecting falls with 3D range camera in ambient assisted living applications: a preliminary study," *Medical Engineering and Physics*, vol. 33, no. 6, pp. 770–781, 2011.
- [16] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings of the 1999*

IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99), pp. 246–252, June 1999.

- [17] M. Isard and A. Blake, “Condensation—conditional density propagation for visual tracking,” *International Journal of Computer Vision*, vol. 29, no. 1, pp. 5–28, 1998.
- [18] C. Won-Seok, K. Yang-Shin, O. Se-Young, and L. Jeihun, “Fast iterative closest point framework for 3D LIDAR data in intelligent vehicle,” in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '12)*, pp. 1029–1034, 2012.
- [19] Y. Xiao, P. Siebert, and N. Werghi, “Topological segmentation of discrete human body shapes in various postures based on geodesic distance,” in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, pp. 131–135, August 2004.
- [20] S. Baloch, H. Krim, I. Kogan, and D. Zenkov, “Rotation invariant topology coding of 2D and 3D objects using morse theory,” in *Proceedings of the IEEE International Conference on Image Processing 2005 (ICIP '05)*, pp. 796–799, September 2005.
- [21] A. Verroust and F. Lazarus, “Extracting skeletal curves from 3D scattered data,” *Visual Computer*, vol. 16, no. 1, pp. 15–25, 2000.
- [22] G. Diraco, A. Leone, and P. Siciliano, “Geodesic-based human posture analysis by using a single 3D TOF camera,” in *Proceedings of the IEEE International Symposium on Industrial Electronics (ISIE '11)*, pp. 1329–1334, 2011.
- [23] W. Brendel and S. Todorovic, “Activities as time series of human postures,” *Lecture Notes in Computer Science*, vol. 6312, no. 2, pp. 721–734, 2010.
- [24] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani, “Probabilistic posture classification for human-behavior analysis,” *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, vol. 35, no. 1, pp. 42–54, 2005.
- [25] R. Debnath, N. Takahide, and H. Takahashi, “A decision based one-against-one method for multi-class support vector machine,” *Pattern Analysis and Applications*, vol. 7, no. 2, pp. 164–175, 2004.
- [26] F. Buccolieri, C. Distanto, and A. Leone, “Human posture recognition using active contours and radial basis function neural network,” in *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS '05)*, pp. 213–218, September 2005.
- [27] S. Park and H. Kautz, “Privacy-preserving recognition of activities in daily living from multi-view silhouettes and RFID-based training,” in *Proceedings of the AAAI Fall Symposium*, pp. 70–77, Arlington, Va, USA, November 2008.
- [28] F. V. Jensen and T. D. Nielsen, “Bayesian networks and decision graphs,” in *Information Science and Statistics*, M. Jordan, J. Kleinberg, and B. Schölkopf, Eds., Springer Science Business Media, New York, NY USA, 2007.
- [29] L. Gond, P. Sayd, T. Chateau, and M. Dhome, “A regression-based approach to recover human pose from voxel data,” in *Proceedings of the IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops '09)*, pp. 1012–1019, October 2009.
- [30] W. Li, Z. Zhang, and Z. Liu, “Action recognition based on a bag of 3D points,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '10)*, pp. 9–14, June 2010.
- [31] “MESA Imaging AG,” SR4000 Data Sheet Rev.5.1., 2011, <http://www.mesa-imaging.ch>.
- [32] A. Ellelt, “Keeping dementia residents safe,” *Assisted Living Consult*, vol. 3, no. 5, pp. 19–41, 2007.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

